



An Offline Corpus for Legal Translations into Italian: a Case Study with a Land Lease Agreement

by Patrizia Giampieri

ABSTRACT: Offline corpora are claimed to be helpful in technical translations. This paper explores whether a corpus of Italian land lease agreement samples and the civil code can be supportive in legal translations. To this end, three experienced translators were involved in a trial lesson and assigned to translate a land lease agreement from English (L2) into Italian (L1). They participated in a webinar where they became acquainted with corpus compilation and analysis. During the training, they clarified linguistic doubts, had the possibility to verify attested usage of words, and they could identify collocational preferences as well as word frequencies. Thanks to the corpus analysis, the translators could improve the quality of their translation work and learn new terminology. The only drawback of using an offline corpus was perceived as lying in its time-consuming compilation and in the fact that in order to be representative, a corpus must be composed of a certain amount of reliable documents.

KEY WORDS: corpus analysis; corpus studies for specialized translations; legal translations; legal texts; corpus-based translations



INTRODUCTION

Although corpus studies for legal translations have been addressed by the literature at length (Orozco-Jutorán and Sánchez-Gijón; Scott, "Genre-specific"; Gallego-Hernández; Vigier Moreno), little research has been carried out on specific case studies revolving around offline corpora (see Vigier Moreno and Sánchez ; Giampieri, *Guida*). Offline corpora, also referred to as DIY (do-it-yourself) (Scott, "Genre-specific"), *ad hoc* (Sánchez-Gijón; Vigier Moreno) and disposable corpora (Varantola, "Disposable"), are collections of texts in electronic format which are compiled by the user for specific purposes or tasks (Jensen, Moustén and Laursen 23). Amongst their features, DIY corpora are claimed to be rather small (Scott, "Genre-specific"). Bowker and Pearson (48) quote that well-designed offline corpora for specialized translations can be of ten thousand to hundreds of thousand words in size. Zanettin corroborates this claim by stating that offline corpora can be composed of a number of texts varying from 10 to 50 (244). The main reasons why offline corpora are rather small lies in the fact that their compilation is extremely time-consuming (244) and sector language does not need large corpora to be attested (Bowker and Pearson 48). The literature reports, in fact, that specialized corpora "can be quite limited and still enjoy terminological representativeness" (Jensen *et al.* 24). The question of representativeness is highly disputed in the literature. As Biel posits, in fact, any claims on statistical frequency or speculations on language attested usage are strictly bound to the language sample collected (Bowker and Pearson 48; Biel 2).

As far as legal corpora are concerned, they tend to be even smaller than any other technical DIY corpora, because legal language is very formulaic and repetitive (Bhatia *et al.* 207; Scott, "Genre-specific" 91). Hence, it is not necessary for the compiler to include many documents to have a representative corpus (Scott, "Genre-specific"). The literature also reports that the case studies dealing with corpora for legal translations are relatively scarce. The main reason is due to the confidentiality of legal documents (Biel 4; Vigier Moreno and Sánchez 261). Therefore, it is often the case that studies on legislation are over-represented (Biel 4; Scott, "Genre-specific" 88), whereas studies on contracts and corporate legal genres are under-represented (Scott, "Towards" 2).

In light of the argumentations above, it is the aim of this paper to present a case study focussing on the compilation and consultation of an offline corpus of land lease agreements. The study was conducted with three experienced translators who were not acquainted with compiling or consulting DIY corpora as language reference tools. In particular, two had more than 9 years' experience in technical translations and one 5. As far as legal translations are concerned, only one had at least 5 years' experience, and the others had 1 year's or no experience at all. Furthermore, two of them had been trained in legal matters.

Before the webinar took place, the participants were asked to fill in a first questionnaire. The questions focused on their experience as (legal) translators; the language resources they used for the translation of the land lease agreement; the time they took to translate it and the difficulties they encountered (if any). The questions and answers to the first questionnaire are reported in Table 1.



Questions	Translator A	Translator B	Translator C
1.Years of experience as a translator	10	9	5
2.Years of experience as a legal translator	5	1	-
3.If you do not have any experience as a legal translator, have you been trained in legal matters?	Yes	Yes, at school and at university (private law, business law, comparative law)	No
4.Language resources you used in your translation (be as specific as possible)	-Wordreference -Reverso	-IATE -Reverso -Linguee -online contract samples -online glossaries	-Proz
5.Which terms/phrases did you find difficult or complicated during the translation?	nothing	-formulaic phrases -legal expressions	one term (PPA)
6.Did you manage to translate the whole text?	Yes	Yes	Yes
7.Time employed to translate	Between 15 and 20 minutes	Between 45 and 60 minutes	120 minutes

Table 1

It is evident that there was one experienced legal translator, one with little and one with no experience in legal translations. It is also remarkable that no dictionaries were mentioned as language resources. On the contrary, the translators seemed to rely more on online language platforms not necessarily implemented by lexicographers (Reverso, Linguee, Wordreference), or on *fora* (Proz). It is also debatable whether the IATE online resource is entirely reliable, as it mirrors the linguistic choices of an online platform (namely, the Eur-Lex) which may not always be linguistically reliable (Giampieri, "Critical" 453-456). Furthermore, as can be noted, only one translator searched for online contract samples and glossaries. The length of the time employed to translate the text reflected the translators' experience, as the least experienced took two hours, whereas the most experienced approximately 20 minutes.

It goes without saying that this paper is mainly aimed at carrying out a qualitative analysis and it is hoped that it will shed light on relevant aspects of corpus-based specialized translations.

ANALYSIS

The participants' first task was to translate a land lease agreement (240 words) from English (L2) into Italian (L1) by using the language resources they generally used. The land lease agreement was aimed at building and operating a power plant on a site. The translators were then invited to participate in a 1-hour webinar where they were explained how to source reliable and representative documents (in Italian) from the Web and how to convert them into text files in order to compile a corpus. To this end, they were introduced to TextSTAT offline concordancer (Hüning).



TRANSLATION SHORTCOMINGS

This paragraph will shed light on the participants' translation shortcomings. The literature generally distinguishes between "mistakes" and "errors". Ellis defines "mistake" as a deviation which occurs when second language (L2) speakers have language processing problems (Ellis 51). An "error", instead, occurs because L2 speakers lack knowledge (51). Other scholars tend to define an error or a mistake as "an unsuccessful bit of language" (James 1). More precisely, they are forms or combinations of forms which would not be produced by native speakers (Lennon 182).

This paper will not distinguish between errors or mistakes, but will refer more generally to translation shortcomings. These will be interpreted as mistranslations or misinterpretation of the source text due to insufficient knowledge of the subject matter or of the legal context. Most likely, translation shortcomings will occur at lexical or collocational level.

Appendix 1 highlights the translation issues, whereas the most recurrent ones will be addressed here below. As stated above, the participants had to submit their translations before the webinar, together with the answers to the first questionnaire. In this way, it was possible to analyse the shortcomings (if any) to tackle during the webinar.

What mostly came to the fore was the rendering of "land lease agreement". All translators, in fact, chose *contratto di locazione di terreno* (literally "land lease/rent contract"). Although the word *locazione* might not seem a mistranslation, it is necessary to clarify that, according to the Italian civil code, a distinction must be made between *locazione* and *affitto*. In the Italian language, *locazione* always refers to buildings and houses, whereas *affitto* to productive goods, such as land and businesses. Unfortunately, the two terms can both be rendered with "lease" and "rent" (Giampieri, *Guida* 17). As a consequence, the fact that online language resources and dictionaries propose *affitto* and *locazione* as translation candidates of "rent" and "lease" often leads to discrepancies, especially amongst non-experienced translators or translators not sufficiently trained in legal matters. It goes without saying that the participants also mistook the contract parties (i.e., the lessor and the lessee) which they wrongly translated *locatore* and *locatario* (deriving from *locazione*), instead of *locatore* and *affittuario* (deriving from *affitto*). All these shortcomings were due to inexperience in the subject matter, but also in the specific legal language.

Other words which were imprecise revolved around acronyms and formulaic expressions. For example, the formulaic expression "by and between" was written at the very beginning of the agreement to designate the two contract parties. This expression can be rendered with *tra* (literally "between") in Italian (see the corpus analysis on the pages below). However, this expression was mistakenly translated, as two participants translated it literally and the third one used a mostly uncommon synonym such as *fra* ("between/among"). These shortcomings were mostly due to lack of knowledge of the legal jargon.



A common issue was the acronym PPA (Power Purchase Agreement), which two translators rendered in Italian (CAE and CAEE respectively), disregarding the fact that the term PPA is also used in Italian documents.

Another mistranslated formulaic expression was “term of lease”, which was the title of a final contract clause. In this respect, the word “term” was considered literally and translated *termine* (meaning “term/condition”), whereas it should have been interpreted “duration”, as the contract clause dealt with the beginning and end of the lease. In this case, the translators' shortcoming was due to a misinterpretation of a word in context.

A last common inaccuracy was a matter of logic: the “production site” involving the land lease agreement was rendered with *ufficio di produzione* or *ufficio produttivo* (literally “production/productive office”). However, no office was foreseen in the contract. Therefore, this term could have been translated literally (i.e., *sito produttivo*), as the “site” in question was the one where the power plant was to be built.

Other shortcomings will be addressed in the corpus analysis paragraph.

CORPUS COMPILATION

After receiving the translations and the first questionnaire, the webinar took place. The webinar was aimed at showing how to compile a corpus on land lease agreements in Italian and how to consult it in order to (easily) translate the land lease agreement assigned. In particular, the corpus analysis focused on tackling the shortcomings described in the paragraph above (and reported in Appendix 1).

For the reasons explained in the paragraph above, before starting the corpus compilation the rendering of “land lease” was addressed. To this end, the Italian civil code was downloaded and it was easily shown that the word *locazione* mostly collocated with houses and lodgings, whereas *affitto* with land. This aspect was crucial and was clarified at the very beginning of the webinar, because the documents which had to be searched on the Internet were obviously related to the words *contratto di affitto*, instead of the wrongly assumed *contratto di locazione*.

Once this aspect was clear, the corpus compilation began. The participants were shown how to look for reliable sources on the Internet. In particular, together with the Italian civil code, land lease contracts (either compiled or as samples) were searched. Therefore, the Google query was written as follows: “*contratto di affitto terreno*” *site:.it*. The documents retrieved (mostly in PDF format) were downloaded, then converted into text files and uploaded on TextSTAT. This mechanism was considered a bit intricate by some participants, especially by those who were not acquainted with file conversion. Also a few Internet sites were considered valid sources. These were uploaded on TextSTAT without converting them into text files. It took approximately 30 minutes to compile a corpus of only 6 texts. The reason was mainly due to the fact that some participants “lost” the downloaded files, could not find where they had saved or converted them, etc. For this reason, after illustrating the principles of corpus compilation, the webinar continued with a corpus which the author had previously compiled. This corpus consisted of 16 files, mainly composed of PDF and DOC files converted into text and a few informative webpages.



CORPUS ANALYSIS

As outlined above, the corpus analysis was aimed at tackling the translators' shortcomings. After showing how to compose the corpus, challenging terms and formulaic expressions were addressed by dividing them into lexical and collocational (i.e., formulaic) instances. Hence, translation candidates of the following words and formulaic expressions were searched in the corpus by generating concordance lines: "principal place of business"; "extension"; "PPA"; "by and between"; "whereas", and "term of lease".

At first, the lexical instances were addressed. As far as the rendering of "principal place of business" was concerned, the word *sede* (literally: "seat"), was searched in the corpus because the participants translated it *sede principale* (literally: "main seat"), *sede operativa principale* (literally "main operative seat"), and *sede legale* (literally: "legal seat"). Many concordances with *sede legale* (literally: "legal seat") and *sede principale* (literally: "main seat") came to the fore, but none with *sede operativa principale* (literally "main operative seat"). This helped clarifying that "principal place of business" could be rendered either with *sede principale* or *sede legale*.

The same strategy was followed with the word "extension", which one participant had translated literally (i.e., *estensione*: "lengthening"). In order to highlight that a literal rendering was not in use in legal language, the word "extension" was searched in the Hoepli online dictionary, in order to find other translation candidates. The dictionary, in fact, proposed *proroga* and *dilazione* (literally: "postponement" and "delay"). In this case, the TextSTAT collocation search function was used and the word *proroga* (literally "postponement") was searched together with *contratto* ("contract") in a span of 5 words on the left and 5 on the right. It was found that the two words collocated in the corpus, whereas *dilazione* ("delay") did not collocate with *contratto* ("contract"). Therefore, the best word to translate "extension (of a contract)" was *proroga*.

As far as the acronym "PPA" was concerned, another strategy was followed. In this case, the string *PPA site*:.it was googled. In this way, the term PPA was searched only in Italian domains. It was remarkable that many Italian documents make use of this English acronym, as 350.000 occurrences were shown. Therefore, PPA could be left as such in the translated text.

The formulaic instances were then tackled. As far as "by and between" was concerned, the translators were shown that the formulaic expression could be rendered with *tra* ("between"). To this end, a literal translation (i.e., *da e tra*) was searched in the corpus. As expected, no concordances were found. Therefore, the preposition *tra* was searched and it was evident that it was followed by words such as *le parti* ("the parties"). Also, *fra* was searched, but very few concordances appeared and they were not related to "the parties".

The participants' proposed translations of "whereas" were *premesso che* (literally "as stated beforehand") and *considerando che* (literally: "considering that"). In order to clarify doubts, "whereas" was searched in the Hoepli online dictionary, which showed *premesso che*, *considerato che* and *visto che* (literally: "as stated beforehand",



“considered that” and “seen that”). The first rendering (*premessso che*) was queried in the corpus and a few concordances came to the fore. The other translation candidates, instead, provided no results. Therefore, it was attested that *premessso che* was an acceptable translation candidate.

Furthermore, in the formulaic expression “term of lease”, the word “term” was mistranslated and rendered literally. As dictionaries provided a literal translation as well, a particular corpus search had to be carried out. In this case, a sort of play-with-word strategy was undertaken and the words which followed “term” were searched in the corpus. Hence, *dell'affitto* (literally: “of the lease”, or “of the rent”) was queried. Interestingly enough, concordances with *durata dell'affitto* (literally: “duration of the lease”) came to the fore, whose content mostly related to the content of the “term of lease” clause.

After tackling the translators' recurrent shortcomings, the training was over and the participants were asked whether it had been useful. Their answers were positive as they felt that corpus responses made them feel more secure about acceptable translation candidates. Furthermore, they were satisfied because they had learnt new terms.

In light of the above, it can be stated that corpus analysis, together with dictionary and Google advanced search, probably helped with the translation of the land lease agreement. In particular, the online bilingual dictionary proposed translation candidates (e.g. of “extension” and “whereas”), which were then queried and found in the corpus. Google advanced search, instead, helped with the translation of highly peculiar terms (such as PPA), which could not be found in the corpus. Sometimes neither dictionaries nor Google advanced search could provide translation candidates (e.g. “term of lease”). In those cases, a sort of “play-with-word” strategy was necessary in order to find concordances (e.g. the back-translation “postponement” as a collocation of “contract” and the back-translation “of the lease” to find the rendering of “term” in the clause title “term of lease”).

THE FINAL QUESTIONNAIRE

After the webinar, the participants were sent a final questionnaire which was aimed at gathering their opinions about the offline corpus and the analysis carried out. The full questionnaire and its answers are reported in Appendix 2. In particular, all participants stated that compiling and using the offline corpus was useful to clarify terms and increase the chances of a well-done translation. Furthermore, they asserted that they would compile offline corpora in the future, especially for terminology-building purposes. One participant only admitted that offline corpus compilation was quite time-consuming; hence, it would be evaluated each time whether a corpus was worth compiling or not.

Another question focused on the perceived weaknesses of an offline corpus. The answers mostly revolved around corpus representativeness, as one participant reported that an obvious hindrance lay in the fact that a corpus can provide answers on the basis of the documents it contains. Another participant remarked that a corpus



must be large to be reliable. The last participant, instead, did not appreciate all the file conversions entailed in the corpus compilation.

Finally, all participants expressed their need for further training in offline corpora. The translators were also asked whether they were acquainted with online corpora. Only one replied positively, but mentioned the Reverso and Linguee online language platforms (see the full final questionnaire in Appendix 2).

DISCUSSION

It is the opinion of the author that the translators were satisfied with the corpus-based translation revision and perceived improvements in translation quality. Most of all, they understood that corpus consultation helps make informed decisions (Vigier Moreno 111) and notice deviant or irrelevant language patterns (De Sutter *et al.* 26). Furthermore, corpus consultation increased their confidence (Varantola, "Translators" 67; Vigier Moreno 104).

It goes without saying that, as remarked by the participants, in-depth training is necessary in order to take full advantages of corpus analysis. However, it is the opinion of the author that training should not only concern corpus consultation, but also compilation. Composing a corpus satisfactorily is crucial for reasons of corpus quality and representativeness (Biel 2). A corpus, in fact, must be representative of its genre or sub-genre (Biel). As highlighted by the literature and as remarked by one participant, a corpus can only tell us what is or is not present in it (Bennet 3). Therefore, the more representative the corpus, the more chances that corpus-based translations will be accurate and reliable.

As far as the corpus compilation and file conversions are concerned, it should be noted that there are programmes which search, convert and compile documents automatically on the basis of keyword input. One of these programmes is BootCAT (Baroni and Bernardini). This programme was not introduced during the webinar because more relevance was given to corpus consultation.

CONCLUSIONS

This paper was aimed at exploring whether an offline corpus composed of 16 texts could be used to translate a legal text. In particular, a land lease agreement had to be translated from English into Italian by three experienced translators with no prior knowledge of corpus compilation and analysis. Two of them had experience in legal translations; the third one had none.

The participants were firstly invited to translate the legal text, then to participate in a 1-hour webinar, where they were explained how to search for reliable documents online and how to compile a corpus. The corpus was analysed by means of TextSTAT concordancer and was composed of the Italian civil code, land lease agreement samples and some informative websites. The corpus analysis was aimed at tackling the translators' shortcomings and providing instances of legal language attested usage. Therefore, the participants learnt how to generate concordances and how to verify



collocational use. In practice, they saw how terms were used in context; they checked the use and frequency of certain formulaic expressions, and how words collocated with others. They understood that certain words or phrases they had proposed were not acceptable translation candidates, as they did not show any occurrences or did not collocate in the corpus. In this way, the participants could clarify linguistic doubts and learn new terms. At the end of the webinar, the participants were administered a questionnaire, where they expressed their opinions about the offline corpus used for the legal translation.

What emerged during the webinar and from the final questionnaire, was a general satisfaction with the corpus analysis and a feeling of trust in the results obtained by corpus interrogation. However, some of the participants did not hide their scepticism about the amount of time spent on corpus compilation and/or about the fact that a wide but careful selection of documents was necessary in order to have a representative and reliable corpus.

In light of the above, the paper findings highlight that offline corpora can be supportive in technical translations, as long as translators are trained and know how to select a sufficiently wide variety of reliable documents from the Web.

WORKS CITED

Baroni, Marco, and Silvia Bernardini. "BootCaT: Bootstrapping corpora and terms from the web." *Proceedings of LREC 2004*, 2004. http://sslmit.unibo.it/~baroni/publications/lrec2004/bootcat_lrec_2004.pdf. Accessed 4 Mar. 2020.

Bennet, Gena R. *Using Corpora in the Language Learning Classroom: Corpus Linguistics for Teachers*. U of Michigan Press, 2010.

Bhatia, Vijay, et al. "Legal discourse: Opportunities and threats for corpus linguistics." *Discourse in the Professions. Perspectives from Corpus Linguistics*, edited by Ulla M. Connor and Thomas A. Upton, John Benjamins, 2004, pp. 203-231.

Biel, Łucja. "Corpus-Based Studies of Legal Language for Translation Purposes: Methodological and Practical Potential." *Reconceptualizing LSP. Online proceedings of the XVII European LSP Symposium 2009*, edited by Carmen Heine and Jan Engberg, 2010, pp. 1-19. <https://pdfs.semanticscholar.org/74a8/ee84fb30d6e5593fba1fa0d1cb02cc3f9030.pdf>. Accessed 4 Mar. 2020.

Bowker, Lynne, and Jennifer Pearson. *Working with Specialized Language*. Routledge, 2002.

De Sutter, Gert, et al. "Towards a corpus-based, statistical approach to translation quality: Measuring and visualizing linguistic deviance in student translation." *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, vol. 16, 2017, pp. 25-39.

Ellis, Rod. *The Study of Second Language Acquisition*. 2nd ed., OUP, 2008.

Eur-lex, <https://eur-lex.europa.eu>. Accessed 4 Mar. 2020.

Gallego-Hernández, Daniel. "The use of corpora as translation resources: A study based on a survey of Spanish professional translators." *Perspectives*, vol. 23, no. 3, 2015, pp. 375-391.



Giampieri, Patrizia. "A Critical Comparative Analysis of Online Tools for Legal Translations." *The Italian Law Journal*, vol. 2, no. 2, 2016, pp. 445-461.

---. *Guida agli strumenti per la traduzione giuridica: corpora offline a supporto del traduttore*. Filodiritto, 2019.

---. "The web as corpus and online corpora for legal translations." *Comparative Legilinguistics*, vol. 33, 2018, pp. 35-55.

Hoepli online bilingual dictionary, <http://dizionari.repubblica.it>. Accessed 4 Mar. 2020.

Hüning, Matthias. TextSTAT (version 2.9), 2008, <http://neon.niederlandistik.fu-berlin.de/en/textstat/>. Accessed 4 Mar. 2020.

IATE (InterActive Terminology for Europe), <http://iate.europa.eu>. Accessed 4 Mar. 2020.

James, Carl. *Errors in Language Learning and Use: Exploring Error Analysis*. Addison-Wesley Longman, 1998.

Jensen, Vigdis *et al.* "Electronic Corpora as Translation Tools: A Solution in Practice." *Communication & Language at Work*, 2012, pp. 21-33.

Lennon, Paul. "Error: Some Problems of Definition, Identification and Distinction." *Applied Linguistics*, vol. 12, no. 2, 1991, pp. 180-196.

Linguee, www.linguee.it. Accessed 4 Mar. 2020.

Orozco-Jutorán, Mariana, and Pilar Sánchez-Gijón. "New Resources for Legal Translators." *Perspectives: Studies in Translatology*, vol. 19, no. 1, 2011, pp. 25-44.

Proz, www.proz.com. Accessed 4 Mar. 2020.

Reverso, www.reverso.net. Accessed 4 Mar. 2020.

Sánchez-Gijón, Pilar. "Developing documentation skills to build do-it-yourself corpora in the specialised translation course." *Corpus use and translating*, edited by Allison Beeby *et al.*, John Benjamins, 2009, pp. 109-127.

Scott, Juliette. "Can Genre-specific DIY Corpora, Compiled by Legal Translators themselves, Assist them in 'Learning the Lingo' of Legal Subgenres?" *Comparative Legilinguistics*, vol. 12, 2012, pp. 87-100. doi.org/10.14746/cl.2012.12.05 Accessed 4 Mar. 2020.

---. "Towards Professional Uptake of DIY Electronic Corpora in Legal Genres." *Salford Working Papers in Translation and Interpreting*, 2012, pp. 1-22.

Varantola, Krista. "Disposable corpora as intelligent tools in translation." *Cadernos de Tradução IX – Tradução e Corpora*, vol. 1, no. 9, 2002, pp. 171-189.

---. "Translators and disposable corpora." *Corpora in Translator Education*, edited by Federico Zanettin *et al.*, St. Jerome, 2003, pp. 55-70.

Vigier Moreno, Francisco, and Sánchez, M María. "Using Parallel Corpora to Study the Translation of Legal System-Bound Terms: The Case of Names of English and Spanish Courts." *Computational and Corpus-Based Phraseology. EUROPHRAS 2017. Lecture Notes in Computer Science*, edited by Ruslan Mitkov, Springer, 2017, pp. 206-273.

Vigier Moreno, Francisco. "Teaching the Use of *ad hoc* Corpora in the Translation of Legal Texts into the Second Language." *Language and Law / Linguagem e Direito*, vol. 3, no. 1, 2016, pp. 100-119.

Wordreference, www.wordreference.com/it/. Accessed 4 Mar. 2020.



Zanettin, Federico. "DIY Corpora: The WWW and the Translator." *Training the Language Services Provider for the New Millennium*, edited by Belinda Maia, Faculdade de Letras, Universidade do Porto, 2002, pp. 239-248.

Patrizia Giampieri has got a Master of Science in Applied Linguistics (Aston University, Birmingham, UK). She lectures English at the School of Law of the University of Camerino, Italy. She lectures legal English in courses for lawyers. She is the author of *Legal English* (Milan, Giuffrè 2015 and 2017) and *Inglese giuridico, Web e corpora (Legal English, Web and Corpora, Milan, Giuffrè 2018)*. She has also written and published articles on legal English and legal translations.

patrizia.giampieri@unicam.it



Appendix 1: Recurrent shortcomings at lexical and collocational level

<i>Phrases (original text)</i>	<i>Incorrect renderings by Translator A</i>	<i>Incorrect renderings by Translator B</i>	<i>Incorrect renderings by Translator C</i>
Principal place of business	—	<i>Sede operativa principale</i> [Back translation: Main operative seat]	—
PPA (Power Purchase Agreement)	—	<i>CAE (Certificato di Acquisto di Energia)</i> [Back translation: Certificate of Purchase of Energy]	<i>CAEE (Certificato di Acquisto di Energia Elettrica)</i> [Back translation: Certificate of Purchase of Electric Energy]
Extension (of the Lease Agreement)	—	—	<i>Estensione</i> [Back translation: extension/lengthening]
Production site	<i>Ufficio produttivo</i> [Back translation: productive office]	<i>Ufficio di produzione</i> [Back translation: production office]	<i>Ufficio di produzione</i> [Back translation: production office]
Incorporated under the laws	—	—	<i>Costituita validamente in base alle leggi</i> [Back translation: validly founded on the basis of the laws]
By and between	<i>Da e tra</i> [Back translation: by and between]	<i>Da e tra</i> [Back translation: by and between]	<i>Fra</i> [Back translation: between/among]
Whereas	—	<i>Considerando che</i> [Back translation: Considering that]	—
Term of Lease	—	<i>Termine della locazione</i> [Back translation: Term/Condition of Lease]	<i>Termini del contratto</i> [Back translation: Terms/Conditions of the contract]



Appendix 2: The Final Questionnaire

<i>Questions</i>	<i>Translator A</i>	<i>Translator B</i>	<i>Translator C</i>
1.The offline corpus helped you (tick as many options as you like): a)clarify some terms b)translate terms I was not able to translate c)increase the chances of a well done translation d)the corpus analysis did not help me at all (say why) e)other (specify)	a, c	a, c e: knowing how to use TextSTAT	a, b, c
2.Do you think you will compile and use an offline corpus for your translations in the future? Why?	Yes, because I will have corpora ready for my most recurrent translations; I will learn terms and clarify doubts.	Yes, because they are very useful for terminology and phraseology search, especially in specialized fields.	Yes, if I have the time, because they help choose the right terms.
3.On the basis of the training, which are the weaknesses of an offline corpus?	None. It goes without saying that queries can be only run on the documents uploaded.	The fact that it must be large to be representative.	The file conversions.
4.Do you feel you would need more training on offline corpora? Why?	Yes, because they are very useful for terminology search and "fine-tuning". Furthermore, I often work offline.	Yes.	Yes, they might be useful in my future translations.
5.Do you know online corpora? Which ones?	No.	Yes, linguee and reverso.	No.
6.If you know online corpora, do you prefer online or offline corpora?	—	I prefer online parallel corpora because they allow to run instant multilingual search.	—